

# Steps Taken for exploring Weather Trends:

## 1. SQL was used to extract data from the source, into two CSV files named as follows:

- a. City\_data.csv
  - i. SQL used: SELECT \* FROM CITY\_DATA;
- b. City\_list.csv
  - i. SQL used: SELECT \* FROM CITY\_LIST;
- c. Global\_data.csv
  - i. SQL used: SELECT \* FROM GLOBAL\_DATA

All the subsequent steps were carried out on Microsoft Excel:

## 2. Filtering out the required data:

As per instruction, my city of residence is Haora, India and I filtered and extracted only those data from the city\_data.csv file with the help of excel.

## 3. Observing data:

- a. In the extracted records for the chosen city, there were records missing between the years: 1808-1812.
  - i. Solutions: A moving average can be calculated for the missing years, to remove the null values from the records.
  - ii. Ignore the null values as a whole

Decision Taken:

much After looking at the data carefully, it can be seen that there is not changes in the data throughout the years. Hence even if a moving average is calculated for the missing records, it won't create any major change in data pattern:

As can be seen from the screenshot below;

1807	Haora	India	25.59	26.164	26.064
1808	Haora	India	26.126667	26.129333	26.120667
1809	Haora	India	26.146389	26.034611	26.120306
1810	Haora	India	26.076088	26.009829	26.118914
1811	Haora	India	26.119095	26.011648	26.182824
1812	Haora	India	26.11652	26.116952	26.140476
1813	Haora	India	25.45	25.981618	26.055476

The records in blue are the ones which are calculated with moving average values from the past records.

Since, these would not create some major change in data pattern, I thought of removing the records as a whole.

## 4. Calculating moving averages:

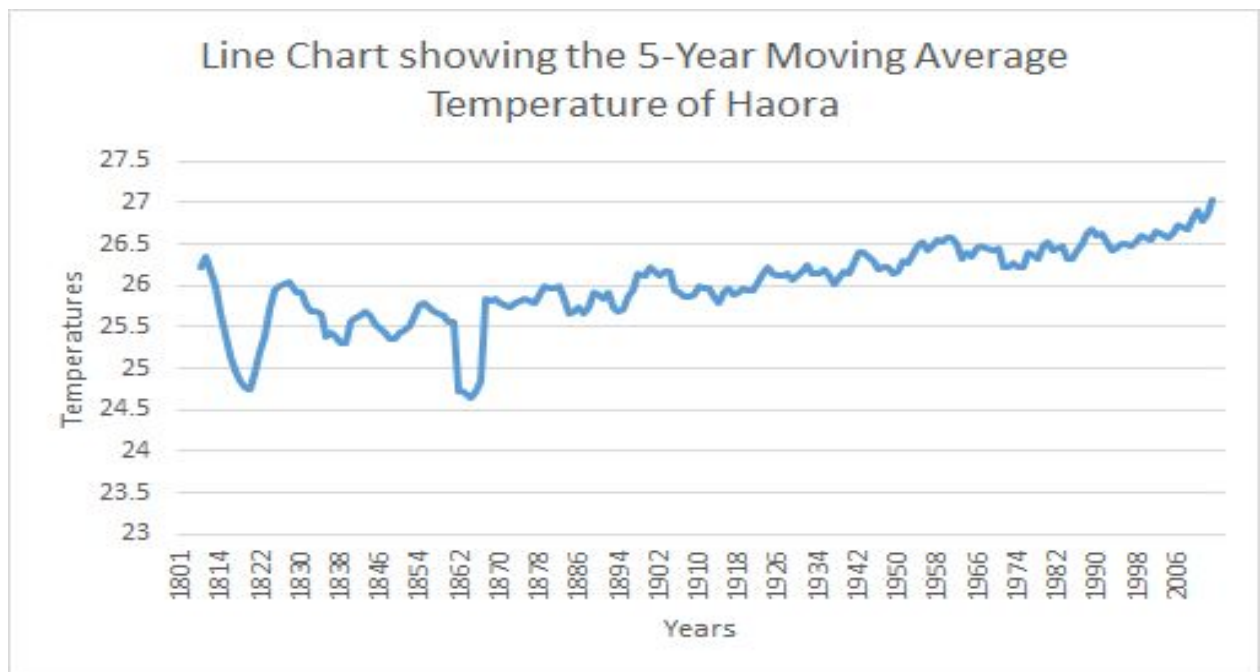
- a. I decided to go for an approach where I was calculating the moving averages from the data of the last 5 years.

**5. Matching the data for a proper Line plot:**

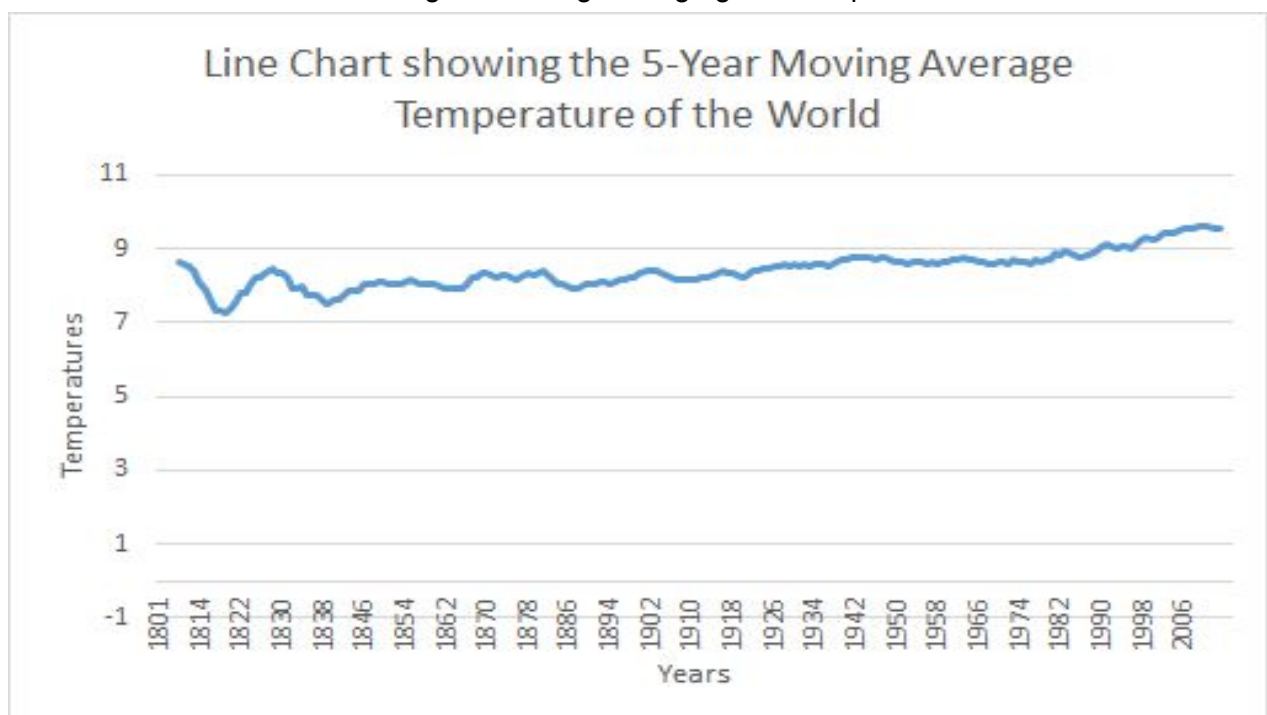
- a. In the global\_data file, there were records for certain years, which were not present for the chosen city, hence I decided to remove those values from the plotting data.

**6. Line Charts and Observations:**

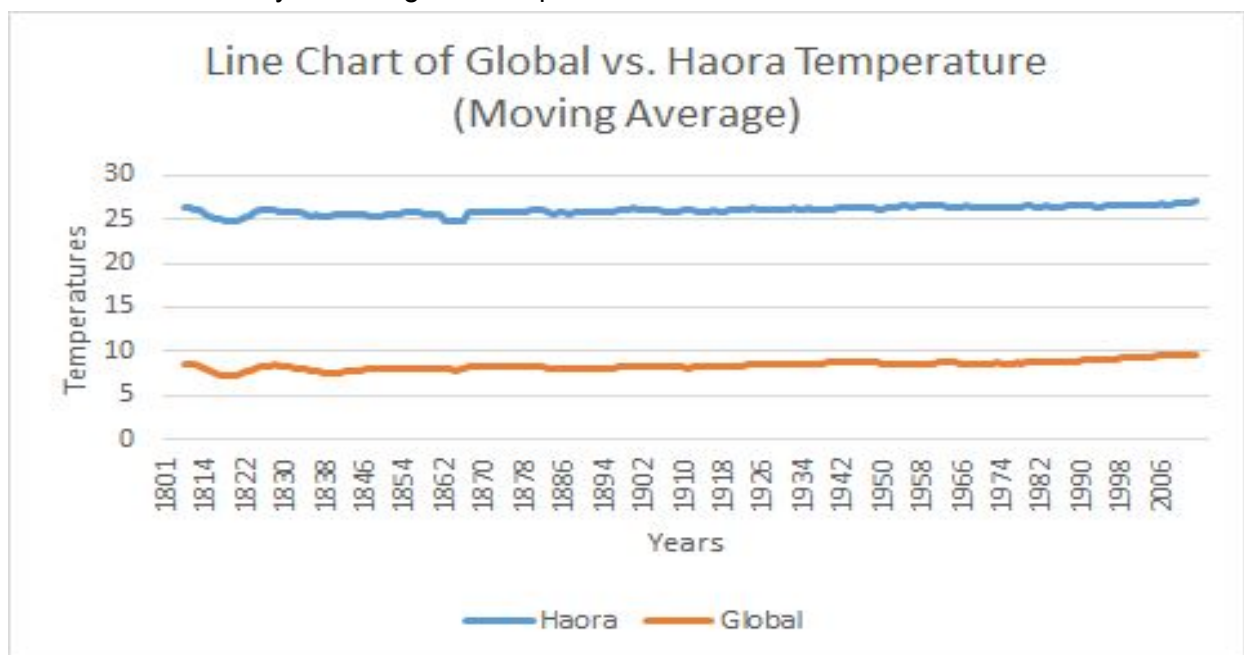
- a. Line chart showing the moving average temperatures of the selected city between the years: 1801-2013



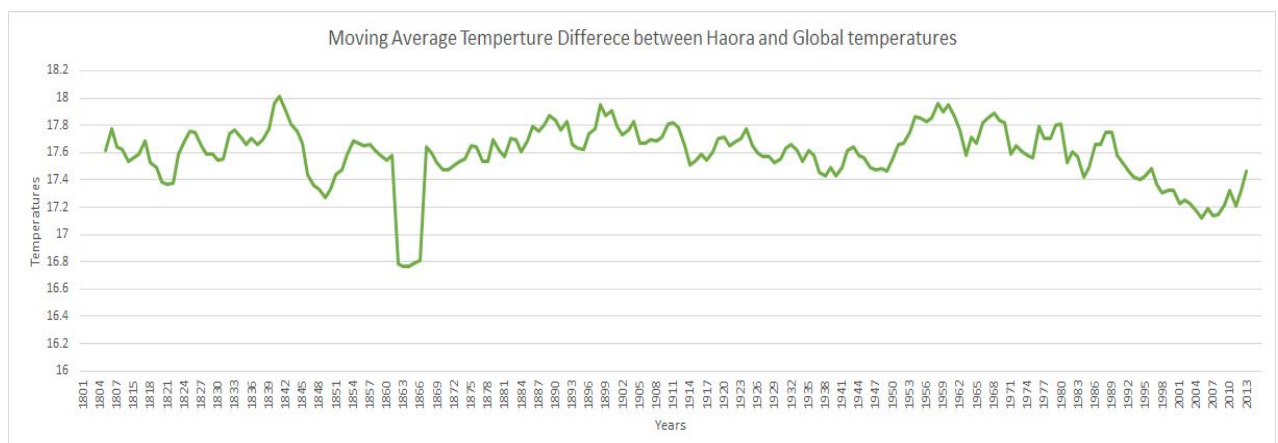
- b. Line chart showing the moving average global temperature:



- c. A combined line chart showing the comparison of the temperatures of both the city and the global temperature:



- d. A final chart showing the temperature differences between the global and the chosen city:



In this chart, I basically calculated the absolute difference between the moving average temperature of the chosen city and the global moving average temperature. I did this mainly to find if there is a major deviation in the average temperatures between the two. If a difference is high, it would indicate that there is a greater difference between the temperature of the chosen city and the global temperature. I might be wrong. Please pardon me if I am wrong.

## Observations:

1. In case of the chosen city we can see there is a significant temperature change (dropping in temperature) in the 15 year time span, between 1806 and 1820. A similar trend can be observed from the global temperature analysis between the years, 1806-1819 (roughly).
2. There were certain changes between the years, 1828-1861, is not as much as the changes in the global average temperature between the years mentioned.
3. Finally as can be seen from the plot of the global average temperatures, there is basically a steady increase in temperature starting from the year 1805 at a recorded temperature of 8.614 to the year 2013 at a temperature recorded at 9.57.
4. For the temperature of our chosen city there were many significant changes between the total range of years (1801-2013) viz. Between the years 1861-1864, there was a change from temperature 25.562 to 24.656, which again increased to 25.84 in 1867.  
(This might be the reason for choosing a moving average step of 5 years.)
5. Finally as can be seen, there is a significant similarity of increasing temperature, in the overall range of years.
  - a. For our chosen city between the years 1805-2013, the temperature increased from 26.228 to 27.032 i.e. a 0.804 rise in overall temperature throughout the years.
  - b. For the global average temperature trend, as already mentioned, there is an increase in temperature of 0.956.
  - c. Although the global temperature seems to be getting a bit steady from 2013 as observed from the line chart plot. In case of the temperature trend for our chosen city, the temperature seems to be on the verge of more increasing in the later years.